

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-278794

(43)Date of publication of application : 22.10.1996

(51)Int.Cl. G10L 3/00
 G10L 3/00
 G10L 3/00
 G06F 17/28

(21)Application number : 07-082218

(71)Applicant : SONY CORP

(22)Date of filing : 07.04.1995

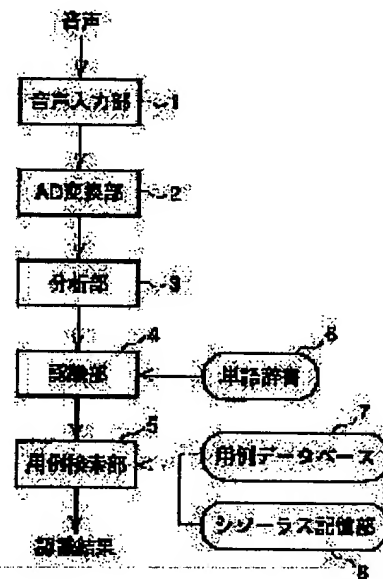
(72)Inventor : ASANO KOJI
 OGAWA HIROAKI
 KATO YASUHIKO
 KAGAMI TETSUYA
 WATARI MASAO
 AKAHA MAKOTO
 ISHII KAZUO
 TANAKA MIYUKI
 TSUNODA HIROSHI

(54) SPEECH RECOGNITION DEVICE AND ITS METHOD AND PHONETIC TRANSLATION DEVICE

(57)Abstract:

PURPOSE: To obtain a speech recognition result without using a grammar rule.

CONSTITUTION: In an analyzing part 3, a voice inputted via a voice input part 1 and an A/D converter 2 is acoustically analyzed and the feature parameter of the voice are extracted. In a recognition part 4. Speech recognitions are performed based on the feature parameters to obtain plural recognition result candidates. Plural examples are stored in an example database 7 and in an example retrieving part 5, degrees of similarities between plural recognition result candidates obtained in the recognition part 4 and examples stored in the example database 7 are calculated and the speech recognition result is obtained based on the degrees of similarities.



LEGAL STATUS

[Date of request for examination] 21.06.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the
 examiner's decision of rejection or application converted
 registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of
 rejection]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-278794

(43) 公開日 平成8年(1996)10月22日

| (51) Int.Cl. ⁸ | 識別記号 | 序内整理番号 | F I | 技術表示箇所 |
|---------------------------|-------|---------|---------------|---------|
| G 1 0 L 3/00 | 5 6 1 | | G 1 0 L 3/00 | 5 6 1 G |
| | 5 5 1 | | | 5 5 1 C |
| | 5 7 1 | | | 5 7 1 C |
| G 0 6 F 17/28 | | 8420-5L | G 0 6 F 15/38 | P |

審査請求 未請求 請求項の数7 O L (全 11 頁)

(21) 出願番号 特願平7-82218

(22) 出願日 平成7年(1995)4月7日

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72) 発明者 浅野 康治

東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内

(72) 発明者 小川 浩明

東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内

(72) 発明者 加藤 靖彦

東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内

(74) 代理人 弁理士 稲本 義雄

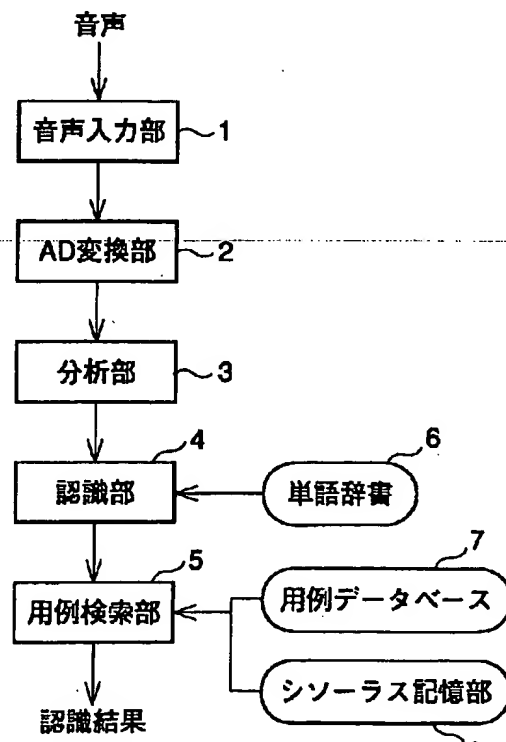
最終頁に続く

(54) 【発明の名称】 音声認識装置および音声認識方法並びに音声翻訳装置

(57) 【要約】

【目的】 文法規則を用いずに、音声認識結果を得ることができるようになる。

【構成】 分析部3では、音声入力部1およびAD変換部2を介して入力された音声の音響分析され、その特徴パラメータが抽出される。認識部4では、その特徴パラメータに基づいて音声の認識が行われ、複数の認識結果候補が求められる。用例データベース7には、複数の用例が記憶されており、用例検索部5では、認識部4で求められた複数の認識結果候補それぞれと、用例データベース7に記憶されている用例それぞれとの類似度が計算され、その類似度に基づいて、音声認識結果が求められる。



【特許請求の範囲】

【請求項1】 文音声認識を行う音声認識装置であって、

音声を音響分析し、その特徴パラメータを抽出する抽出手段と、

前記抽出手段より出力される前記特徴パラメータに基づいて前記音声の認識を行い、複数の認識結果候補を求める認識手段と、

複数の用例を記憶している用例記憶手段と、

前記複数の認識結果候補それぞれと、前記用例記憶手段に記憶されている用例それぞれとの類似度を計算し、その類似度に基づいて、前記音声の認識結果を求める計算手段とを備えることを特徴とする音声認識装置。

【請求項2】 単語を、その概念ごとに分類して記憶している単語概念記憶手段をさらに備え、

前記認識結果候補は、1以上の単語の組み合わせであり、

前記計算手段は、前記認識結果候補を構成する単語それぞれと、前記用例を構成する単語それぞれとの概念的な類似性を表す単語類似度を、前記単語概念記憶手段を参照して求め、その単語類似度に基づいて、前記類似度を計算することを特徴とする請求項1に記載の音声認識装置。

【請求項3】 前記単語概念記憶手段は、単語を、その概念に基づいて木構造に階層化して記憶していることを特徴とする請求項2に記載の音声認識装置。

【請求項4】 言語モデルを記憶している言語モデル記憶手段をさらに備え、

前記認識手段は、前記言語モデル記憶手段に記憶されている言語モデルの制約の下、前記音声の認識を行い、複数の認識結果候補を求めることを特徴とする請求項1に記載の音声認識装置。

【請求項5】 前記認識手段による認識の対象となる単語を記憶している単語記憶手段をさらに備え、前記単語概念記憶手段および単語記憶手段は、一体化されていることを特徴とする請求項2に記載の音声認識装置。

【請求項6】 複数の用例を記憶している用例記憶手段を備え、文音声認識を行う音声認識装置の音声認識方法であって、

音声を音響分析し、その特徴パラメータを抽出し、前記特徴パラメータに基づいて前記音声の認識を行い、複数の認識結果候補を求め、前記複数の認識結果候補それぞれと、前記用例記憶手段に記憶されている用例それぞれとの類似度を計算し、その類似度に基づいて、前記音声の認識結果を求めることを特徴とする音声認識方法。

【請求項7】 入力された音声の言語を、他の言語に翻訳する音声翻訳装置であって、

抽出手段と、

前記抽出手段より出力される前記特徴パラメータに基づいて前記音声の認識を行い、複数の認識結果候補を求める認識手段と、

複数の用例を記憶している用例記憶手段と、

前記複数の認識結果候補それぞれと、前記用例記憶手段に記憶されている用例それぞれとの類似度を計算し、その類似度に基づいて、前記音声の認識結果を求める計算手段と、

10 前記計算手段より出力される前記音声の音声認識結果の言語を、前記他の言語に翻訳する翻訳手段とを備えることを特徴とする音声翻訳装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、例えば音声を認識し、その認識結果の言語を他の言語に翻訳する場合などに用いて好適な音声認識装置および音声認識方法並びに音声翻訳装置に関する。

【0002】

20 【従来の技術】図10は、従来の、文音声認識（会話音声認識、あるいは音声理解）を行う音声認識装置の一例の構成を示している。音声入力部1は、そこに入力される、いわば空気の波である音声を、電気的な波である音声信号に変換する、例えばマイクなどと、そのマイクから出力される音声信号を増幅するアンプなどで構成されている。AD変換部2は、音声入力部1から出力されるアナログの音声信号を、所定のクロックのタイミングでサンプリングし、量子化を行い、デジタルの音声信号（デジタルデータ）に変換するようになっている。

30 【0003】分析部3は、AD変換部2より出力される音声信号を音響分析し、これにより、例えば所定の帯域ごとの音声のパワーや、線形予測係数（LPC）、ケプストラム係数などの音声の特徴パラメータを抽出するようになっている。即ち、分析部3は、例えばフィルタバンクにより、音声信号を所定の帯域ごとにフィルタリングし、そのフィルタリング結果を整流平滑化することで、所定の帯域ごとの音声のパワーを求めるようになっている。あるいは、分析部3は、入力された音声に対し、例えば線形予測分析処理を施すことで、線形予測係数を求め、またその線形予測係数からケプストラム係数を求めるようになっている。

40 【0004】分析部3で求められた特徴パラメータは、そのまま、あるいは、そこで必要に応じてベクトル量子化されて、認識部50に出力されるようになっている。

【0005】認識部50は、分析部3からの特徴パラメータ（あるいは、特徴パラメータをベクトル量子化して得られるシンボル）に基づき、例えばダイナミックプログラミング（DP）マッチング法や、隠れマルコフモデル

音声認識を行うようになされている。ここで、認識部50において、音韻単位での音声認識が行われるとした場合、その結果得られる音声認識結果候補としての音韻の候補（音韻候補）が、例えばラティス（格子形データ）の形で、パーザ部51に出力される（このラティスの形で出力される音韻候補を、以下、適宜、音韻候補ラティスという）。

【0006】パーザ部51は、まず、単語（あるいは形態素）の見出し（例えば、読み出すという単語であれば「読み出す」という見出し）、その音韻情報（読み）、その品詞、およびその他の必要な情報が登録されている単語辞書52を参照し、認識部50からの音韻候補ラティスに含まれる音韻候補を組み合わせて1以上の単語とし、さらにその単語を1以上組み合わせて単語列（文）を作成する。そして、この単語列に対し、所定のパーザングアルゴリズムに基づいて、文法規則辞書53に登録されている文法規則を適用する。パーザ部51は、認識部50からの音韻候補ラティスから作成可能な単語列について、文法規則を適用することを繰り返し、文法規則に最も合致する単語列、即ち文を、文音声認識結果（あるいは音声理解結果）として出力するようになされている。なお、文法規則は、例えば音声認識の分野などにおける専門家によって作成される。

【0007】以上のように構成される音声認識装置では、音声入力部1に音声が入力されると、その音声は、音声信号として、AD変換部2を介して分析部3に出力される。分析部3では、音声信号が音響分析され、音声の特徴パラメータが抽出される。この特徴パラメータは、認識部50に供給され、認識部50では、その特徴パラメータを用いて音素単位での音声認識が行われ、その結果得られる音韻候補ラティスが、パーザ部51に出力される。パーザ部51では、単語辞書52および文法規則辞書53を参照して、認識部50からの音韻候補ラティスから得られる単語列が、文音声認識（音声理解）され、その認識結果が出力される。

【0008】ところで、パーザ部51で文音声認識を行うためのパーザングアルゴリズムとしては、例えば文脈自由文法に基づき、解析中に複数の可能性がある場合には全ての可能性を試し、途中結果を残しながら並列的に解析を進める方法（このような方法によりパーザングを行うパーザは、チャートパーザと呼ばれる）や、文脈自由文法から解析に必要な情報をあらかじめ計算してLR解析表と呼ばれる表を自動的に作成し、それを参照しながら解析を進める方法（このような方法によりパーザングを行うパーザは、拡張LRパーザと呼ばれる）などが知られている。

【0009】ここで、上述のLRパーザのLRは、次のことを意味する。即ち、LRのうちのLは、Left-to-rightの略で、文を左から右方向に走査することを意味する。また、Rは、Right-to-leftの略で、文を右から左方向に走査することを意味する。

erivation（最右導出）の略で、文脈自由文法の書換規則において、最も右の非終端記号に規則を適用することを意味する。なお、文を左から右方向に走査するとともに、最右導出の逆の最左導出を行うパーザもあり、このようなパーザは、LLパーザと呼ばれる。

【0010】また、例えば「HMM音韻認識と拡張LR構文解析法を用いた連続音声認識」、北 他、情報処理学会論文誌、Vol.31, No.3, pp.472-480 (1990)や、

「拡張LR構文解析法を用いた連続音声認識」、伊藤 他、電子情報通信学会技術報告、SP90-74 などでは、上述の方法を改良したものが提案されている。即ち、これらでは、パーザ部において、拡張LRパーザが生起する音韻列を予測し、予測された音韻に対して認識部においてHMMによって音韻照合を行う方法が提案されており、この方法によれば、認識処理における計算量を削減することができる。

【0011】さらに、「確率文法を使った音声言語処理」、北、人工知能学会 言語・音声理解と対話処理研究会、SIG-SLUD-9204-6では、拡張LRパーザにおいて適用する文法規則に、規則の適用確率の情報を考慮する方法が提案されている。即ち、この方法では、まず文法規則として用いる文脈自由文法に対して、大量の学習用テキストを用いて各生成規則の適用確率を求め、その文脈自由文法をLR解析表に展開する際にLR解析表に確率情報も表現するようにし、拡張LRパーザによる解析の途中で生起確率の計算を行うようになされている。これにより、出現頻度の高い言語表現には高い確率値が得られ、非文（文法的に誤っている文）には非常に小さな確率値が得られるようになされている。この方法によれば、さらなる計算量の削減を図ることができる。

【0012】一方、以上のような方法に対して、最初に計算的に負荷の少ない情報を用いて、複数の仮説を求め、その仮説を、より高次の情報で絞り込む、いわば言語的な制約を2段階に分けて行う方法が、例えば“A Trellis Based Fast Search for Finding the N Best Sentence Hypotheses in Continuous Speech Recognition,” F.K. Soong and et.al., Proceedings of Speech and Natural Language Workshop, DARPA, pp.12-19, (1990).や、“The N-best algorithm: An efficient and exact procedure for finding the N most likely sentence hypotheses,” R. Schwartz and et.al., Proceedings of ICASSP 90, pp.81-84 (1990).などに記載されている。

【0013】この方法では、まず認識部において、分析部の分析結果を用いて、例えばHMMなどにより認識処理を行う際に、例えばバイグラム（Bigram）、トリグラム（Trigram）などの統計的言語モデルや有限状態ネットワークなどを用いて緩い言語的制約をかける。バイグラム、トリグラムは、例えば1次、2次のマルコフ過程のモデルで、文の各単語の出現確率が、前単語のみに依存する。また、N-gramは、N-1次のマルコフ過程のモデルで、文の各単語の出現確率が、前N-1単語のみに依存する。

テキストデータベースを基にして学習したものであり、自然言語の局所的な性質を精度良く近似することのできるモデルとして知られている。

【0014】そして、このような制約下において、例えば公知であるビタビ (Viterbi) アルゴリズムを用いたビームサーチなどにより、適当に枝刈を行い、その結果残った複数の文候補を出力する。その後、その複数の文候補に対して、より精細な文法規則を適用して、文音声認識 (音声理解) を行う。

【0015】

【発明が解決しようとする課題】ところで、上述した従来の方法において、パーザ部で用いる文法規則は、人間の話す様々な文の形態を網羅するように専門家が作成しているが、この規則を、適切に (正確な文音声認識を行うことができるように) 記述することは非常に困難であった。即ち、文法規則を厳密にし過ぎると、その規則に合致しない発話は、すべて文として認識されず、逆に、文法規則を緩くし過ぎると、文として構文的に意味をなさない (文法的に誤った) ものが認識結果とされる課題があった。

【0016】そこで、パーズングが適切になされない文例 (発話) を見つけ、その文例に基づいて、文法規則を改善する方法がある。しかしながら、文法規則を、どのように修正するのが効果的かは、一般的に分かりにくく、文法規則を変更することによって、認識処理がどのように変化するかを予測するのが困難であったため、変更前までは、正しい処理がなされていた文に対して、変更後は、誤った処理がなされる場合があり、文法規則の修正は容易ではなかった。

【0017】本発明は、このような状況に鑑みてなされたものであり、文法規則を用いずに、発話の音声認識結果候補の文法的適格性を判定することができるようにし、これにより精度の高い文音声認識 (音声理解) を、容易に行うことができるようにするものである。

【0018】

【課題を解決するための手段】本発明の音声認識装置は、複数の用例を記憶している用例記憶手段と、複数の認識結果候補それぞれと、用例記憶手段に記憶されている用例それぞれとの類似度を計算し、その類似度に基づいて、音声の認識結果を求める計算手段とを備えることを特徴とする。

【0019】本発明の音声認識方法は、複数の用例を記憶している用例記憶手段を備え、文音声認識を行う音声認識装置の音声認識方法であって、複数の認識結果候補それぞれと、用例記憶手段に記憶されている用例それぞれとの類似度を計算し、その類似度に基づいて、音声の認識結果を求めることを特徴とする。

【0020】本発明の音声翻訳装置は、複数の用例を記憶している用例記憶手段と、複数の認識結果候補それぞれと、用例記憶手段に記憶されている用例それぞれとの

類似度を計算し、その類似度に基づいて、音声の認識結果を求める計算手段と、計算手段より出力される音声の音声認識結果の言語を、他の言語に翻訳する翻訳手段とを備えることを特徴とする。

【0021】

【作用】本発明の音声認識装置においては、用例記憶手段は、複数の用例を記憶しており、計算手段は、複数の認識結果候補それぞれと、用例記憶手段に記憶されている用例それぞれとの類似度を計算し、その類似度に基づいて、音声の認識結果を求めるようになされている。

【0022】本発明の音声認識方法においては、複数の認識結果候補それぞれと、用例記憶手段に記憶されている用例それぞれとの類似度が計算され、その類似度に基づいて、音声の認識結果が求められるようになされている。

【0023】本発明の音声翻訳装置においては、用例記憶手段は、複数の用例を記憶している。計算手段は、複数の認識結果候補それぞれと、用例記憶手段に記憶されている用例それぞれとの類似度を計算し、その類似度に基づいて、音声の認識結果を求め、翻訳手段は、計算手段より出力される音声の音声認識結果の言語を、他の言語に翻訳するようになされている。

【0024】

【実施例】図1は、本発明を適用した音声認識装置の第1実施例の構成を示している。なお、図中、図10における場合と対応する部分については、同一の符号を付し、以下、その説明は、適宜省略する。

【0025】認識部4 (認識手段) は、図10の認識部50と同様に、分析部3からの特徴パラメータ (あるいは、その特徴パラメータをベクトル量子化したシンボル) を用いて、所定の音声認識アルゴリズムにしたがい、音声認識を行うようになされている。但し、認識部50では、音素単位の音声認識が行われるようになされていたが、認識部4では、例えば後述する単語辞書6を参照して、単語単位の音声認識が行われるようになされている。そして、認識部4は、単語単位の音声認識の結果得られる音声認識結果候補としての単語の候補 (単語候補) が、例えばラティスの形で出力されるようになされている (このラティスの形で出力される単語候補を、以下、適宜、単語候補ラティスという)。

【0026】用例検索部5 (計算手段) は、認識部4からの単語候補ラティスに含まれる単語候補を組み合わせて、少なくとも1以上の単語からなる単語列 (文) を、複数作成する。そして、その複数の単語列 (文) それぞれと、後述する用例データベース7に記憶されている用例それぞれとの類似度を計算し、その類似度に基づいて、音声入力部1に入力された音声の認識結果 (文) を決定するようになされている。

【0027】なお、用例検索部5における類似度の計算は、認識部4から出力される単語候補ラティスに含まれる単語候補を、単語列 (文) として抽出し、用例データベース7に記憶されている用例それぞれとの類似度を計算する。

用いて行われるようになされている。また、単語候補ラティスに含まれる単語候補を組み合わせ、少なくとも1以上の単語からなる単語列(文)を、複数作成する処理は、用例検索部5ではなく、認識部4に行わせるようにすることができる。

【0028】単語辞書6(単語記憶手段)には、単語の見出しおよびその音韻情報(読み)、さらに必要ならば単語の品詞その他の情報が対応付けられて記憶(登録)されている。認識部4では、この単語辞書6に記憶されている単語を対象として、音声認識が行われるようにな

されている。

【0029】用例データベース7(用例記憶手段)には、複数の用例が記憶(登録)されている。この用例は、例えば新聞に掲載されている文章や、アナウンスが読み上げる原稿に記載の文章などなどに基づいて作成される。

【0030】シソーラス記憶部8(単語概念記憶手段)は、少なくとも単語辞書6に登録されている単語を、その概念ごとに分類して記憶している。本実施例では、シソーラス記憶部8には、後述する図5に示すように、単語を、その概念に基づいて木構造に階層化したシソーラスが記憶されている。

【0031】以上のように構成される音声認識装置においては、音声入力部1に入力された音声は、AD変換部2、さらには分析部3(抽出手段)を介することにより特徴パラメータ(あるいはシンボル)とされて、認識部4に出力される。認識部4では、分析部3の出力を用いて、例えばHMMにしたがい、単語単位の音声認識が行われる。

【0032】ここで、HMMについて簡単に説明する。HMMは、非決定有限状態オートマトンとして定義され、そのモデルは、幾つかの状態と、その状態間の遷移を表すパスから構成される。このようなモデルにおいて、各状態からの状態の遷移過程はマルコフ過程とされ、また、状態が遷移するときにはシンボルが1つ出力されるものとして、モデルの学習が行われる。いま、モデルの状態がN個あり、モデルから出力されるシンボルの種類がK個あるとすると、この学習では、多数の学習データを用い、状態が、状態iから状態jに遷移する確率(状態遷移確率) a_{ij} と、そのときにシンボル y_k が出力される確率(出力シンボル確率) $b_{ij}(y_k)$ が求められる(但し、 $0 < i, j < N+1$, $0 < k < K+1$)。

【0033】なお、HMMのパラメータには、最初に状態iにいる確率(初期状態確率) π_i もあるが、音声認識では、状態が、自分自身か、あるいは自身より右側の状態にしか遷移しないleft-to-rightモデルが、通常用いられるので、初期状態は、モデルの最も左側の状態とされる(最初に、最も左側の状態にいる確率が1とされ、他の状態は0とされる)。

常は、学習において、初期状態確率を求める必要はない。

【0034】一方、認識時には、学習の結果得られた状態遷移確率および出力シンボル確率を用いて、分析部3から出力されるシンボル系列が観測(生起)される確率(生起確率)が計算され、その確率の高いものが認識結果とされる。

【0035】本実施例では、認識部4には、あらかじめ学習を行うことにより得られた、例えば音素単位のモデル(音素モデル)が記憶されており、認識部4は、単語辞書6に登録されている単語の音韻情報を参照して、音素モデルを連結し、単語辞書6に登録されている単語のモデルを作成する。そして、このモデルを用いて、上述したように生起確率を計算し、その確率の高い単語からなるラティスを、単語候補ラティスとして出力する。

【0036】なお、認識部4には、音素モデルではなく、例えば単語単位のモデル(単語モデル)を記憶させておき、そのモデルをそのまま用いて、連続音声認識させるようにすることも可能である。

【0037】認識部4から出力された単語候補ラティスは、用例検索部5に供給される。用例検索部5は、単語候補ラティスを受信すると、例えば図2のフローチャートにしたがった処理を行う。即ち、まず最初に、ステップS1において、単語ラティスを構成する単語が組み合わせられ、少なくとも1以上の単語からなる単語列(文)が作成される。なお、このとき、単語候補ラティスの単語は、時間軸方向に重なりを生じないように、かつ時系列に組み合わせられる。

【0038】即ち、例えば、いま、図3に示すような、時刻t1乃至t5を音声区間とする単語候補ラティスが、認識部4から出力されたとする。なお、図3では、時刻t1乃至t2, t2乃至t4, t4乃至t5, t1乃至t3, t3乃至t5において、単語a, b, c, d, eが、それぞれ生起確率の最も高い単語として認識された様子を示している。

【0039】この場合、時間軸方向に重なりを生じない、時系列の単語の組み合わせは、(a, b, c), (d, e), (a, e), (d, c)の4通りとなる。

【0040】ステップS1では、上述のうちのいずれかの組み合わせが作成される。

【0041】なお、実際には、認識時に音韻のセグメンテーションに誤差が生じるので、図3に示したように、ある単語の音声区間の直後に、他の単語の音声区間が位置することはほとんどなく、連続に発話された単語A, B, Cであっても、通常は、例えば図4に示すように、単語AとBとの音声区間は時間的に重なりを生じ、また単語BとCとの音声区間は時間的に離れたものとなる。そこで、ステップS1では、時間軸方向に重なりを生じないよ

りは許容して、単語の組み合わせが作成されるようになされている。

【0042】単語の組み合わせが作成された後は、ステップS2に進み、その単語の組み合わせと、用例データベース7に登録されている用例それぞれとの類似度が計算される。ここで、本実施例では、この類似度を計算する方法として、例えば特開平3-276367号に開示されているような、単語を、その意味の類似性（概念）に基づいて木構造に階層化したシソーラスを用いる方法を適用する。即ち、単語の組み合わせを構成する、ある単語と、その単語に対応する、用例を構成する単語とが、同一のカテゴリに属すると考えられる概念の階層が第k階層であった場合に、これらの単語間の概念的な類似性を表す単語類似度を、 $(k-1)/n$ （但し、nは、シソーラスの階層数）とし、単語の組み合わせを構成する単語それぞれと、用例を構成する単語それぞれについての単語類似度を積算する。そして、その積算結果を、単語の組み合わせと、用例との類似度とする。

【0043】具体的には、例えば、いま、シソーラス記憶部8に、図5に示すようなシソーラスが記憶されているものとした場合、以下のようにして類似度が計算される。

【0044】但し、図5においては、長方形で囲んであるものは概念を表し、楕円で囲んであるものが単語を表す。同図では、最も上の階層（第4階層）に属する概念が、「変動」、「行動」、「人物」、「社会」、その他に分類されており、そのうちの、例えば概念「人物」は、それに含まれる概念「人称」、「親族」、その他に分類されている。さらに、概念「人称」は、それに含まれる概念「自称」、「他称」、その他に分類されており、そのうちの、例えば概念「他称」には、その範疇にある単語「彼」、「彼女」、その他が属するものとされている。

【0045】また、図5において、最も下の概念の階層を第1階層とし、下から2番目の概念の階層を第2階層とし、以下同様にして、下から3番目の概念の階層、または最も上の概念の階層を、それぞれ第3階層、または第4階層とする。図5のシソーラスは4階層で構成されるから、シソーラスを第1階層までさかのぼることにより概念が一致する単語どうしの単語類似度は0（ $= (1-1)/4$ ）となり、また、シソーラスを第2階層までさかのぼることにより概念が一致する単語どうしの類似度は $1/4$ （ $= (2-1)/4$ ）となる。以下同様に、シソーラスを第3または第4階層までさかのぼることにより概念が一致する単語どうしの単語類似度は $1/2$ または $3/4$ となる。

【0046】例えば、いま、単語辞書6には、自立語のみが登録されており、従って認識部4では、自立語のみを対象として連続音声認識が行われ、これにより、ステップS1で単語「彼」、「任地」、「赴く」の組み合わせ

せ（「彼」、「任地」、「赴く」）（以下、適宜、入力単語列という）が作成されたとするとともに、用例として、例えば「私は学校に行く」を考えた場合、まず、入力単語列（「彼」、「任地」、「赴く」）を構成する単語「彼」、「任地」、「赴く」それぞれと、それぞれと、用例「私は学校に行く」を構成する、「彼」、「任地」、「赴く」に対応する単語「私」、「学校」、「行く」それぞれとの単語類似度は、次のようになる。ここで、単語XとYとの単語類似度を、 $d(X, Y)$ と表す。

【0047】即ち、単語「彼」と「私」とは、第2階層までさかのぼることにより概念「人称」に一致するので、単語類似度 d （「彼」、「私」）は $1/4$ となる。また、単語「任地」と「学校」とは、第3階層までさかのぼることにより概念「社会」に一致するので、単語類似度 d （「任地」、「学校」）は $1/2$ となる。さらに、単語「赴く」と「行く」とは、第1階層までさかのぼることにより概念「往復」に一致するので、単語類似度 d （「赴く」、「行く」）は0となる。

【0048】以上の単語類似度を積算すると、その積算値は $3/4$ （ $= 1/4 + 1/2 + 0$ ）となり、これが、入力単語列（「彼」、「任地」、「赴く」）と用例「私は学校に行く」との類似度とされる。

【0049】以上のようにして、入力単語列（「彼」、「任地」、「赴く」）に対する類似度の計算が、用例データベース7に登録されているすべての用例について行われる。

【0050】図2に戻り、ステップS2で類似度の計算が終了すると、ステップS3に進み、認識部4からの単語候補ラティスから得られるすべての単語の組み合わせについて、ステップS2の類似度の計算を行ったかが否かが判定される。ステップS3において、単語候補ラティスから得られるすべての単語の組み合わせについて、類似度の計算を行っていないと判定された場合、ステップS1に戻り、新たな単語の組み合わせ（入力単語列）を作成し、以下、同様の処理を繰り返す。

【0051】即ち、これにより、ステップS1において、例えば新たな入力単語列として、例えば（「彼女」、「母」、「似ている」）が作成された場合には、この入力単語列を構成する単語「彼女」、「母」、「似ている」（但し、「似ている」は「似る」とされる）それぞれと、上述した用例「私は学校に行く」を構成する、「彼女」、「母」、「似ている」に対応する単語「私」、「学校」、「行く」それぞれとの単語類似度 d （「彼女」、「私」）、 d （「母」、「学校」）、 d （「似る」、「行く」）は、ステップS2において上述した場合と同様にして、 $1/4$ 、 $3/4$ 、 $3/4$ と計算され、その結果、入力単語列（「彼女」、「母」、「似ている」）と用例「私は学校に行く」との類似度は $7/4$

【0052】一方、ステップS3において、単語候補ラティスから得られるすべての単語の組み合わせについて、類似度の計算を行ったと判定された場合、ステップS4に進み、類似度が最も高い用例と入力単語列とが選択され、ステップS5に進む。ステップS5では、ステップS4で選択された用例を構成する単語のうち、同じくステップS4で選択された入力単語列を構成する単語に対応するものが、その入力単語列を構成する単語に、それぞれ置き換えられ、それが、文音声認識結果として出力されて、処理を終了する。

【0053】即ち、例えば、いま、説明を簡単にするために、用例として「私は学校に行く」のみが用例データベース7に記憶されており、入力単語列として、

(「彼」、「任地」、「赴く」)および(「彼女」、「母」、「似ている」)の2つだけが作成されたとすると、上述したように、入力単語列(「彼」、「任地」、「赴く」)または(「彼女」、「母」、「似ている」)それぞれと、用例「私は学校に行く」の類似度は、3/4または7/4となる。本実施例では、類似度が高い場合というのは、類似度の値が小さい場合であり(これは、図5において、シソーラスを構成する最も下の概念の階層から、第1階層、第2階層、・・・としたため、これとは逆に、シソーラスを構成する最も上の概念の階層から、第1階層、第2階層、・・・とすれば、類似度が高い場合というのは、類似度の値が大きい場合となる)、従って、ステップS4では、入力単語列(「彼」、「任地」、「赴く」)と用例「私は学校に行く」とが選択される。

【0054】そして、ステップS5では、用例「私は学校に行く」を構成する単語のうち、入力単語列(「彼」、「任地」、「赴く」)を構成する単語に対応するもの、即ち「彼」、「任地」、「赴く」に対応する「私」、「学校」、「行く」が、その入力単語列を構成する単語「彼」、「任地」、「赴く」に、それぞれ置き換えられる。そして、その置き換え結果「彼は任地に赴く」が文音声認識結果として出力される。

【0055】以上のように、この音声認識装置によれば、音声認識結果候補として複数の入力単語列が出力された場合に、それらが、いわば用例によって絞り込まれ、用例に最も類似する入力単語列が音声認識結果として得られる。従って、複数の音声認識結果候補を絞り込む(複数の音声認識結果候補から、1つの音声認識結果を選択する)のに、専門家が記述した文法規則を用いる必要がなく、また、用例は、新聞等の記載から容易に作成することができるので、装置の開発(製作)を容易に行うことが可能となる。

【0056】さらに、正しい音声認識結果が得られない入力音声があった場合には、その音声を、用例として、用例データベース7に追加するだけで済み、認識性能を容易に改善することができ、その場合、例えば、

S2で類似度を計算する対象となる用例が増えるだけであるから、用例の追加前まで正しく認識されていた音声も、用例の追加後に、正しく認識されなくなるようなこともない。

【0057】また、用例データベース7に、用例とともに、その意味表現などを対応付けて登録しておくようにすることで、入力された音声の意味内容を理解することが、容易に可能となる。

【0058】次に、図6は、本発明を適用した音声認識装置の第2実施例の構成を示している。なお、図中、図1における場合と対応する部分については、同一の符号を付してある。即ち、この音声認識装置は、認識部4に代えて認識部11(認識手段)が設けられ、さらに言語モデル記憶部12(言語モデル記憶手段)が新たに設けられている他は、図1の音声認識装置と同様に構成されている。

【0059】言語モデル記憶部12は、例えばバイグラムやトリグラムなどの統計的言語モデルなどを記憶しており、認識部11は、例えば認識部4における場合と同様の音声認識処理を、言語モデル記憶部12に記憶されている言語モデルにより緩い言語的制約をかけ、その制約の下、例えばビタビアルゴリズムを用いたビームサーチなどにより、適当に絞りながら行う。そして、その結果残った、複数の音声認識結果としての複数の文候補を得て、各文候補を、そこから、例えば助詞を削除することにより自立語のみでなる単語列に変換して、用例検索部5に出力する。

【0060】用例検索部5では、認識部11からの複数の単語列を、入力単語列として、用例との類似度が計算される。

【0061】従って、認識部11では、言語モデルによる制約が緩いため、そこから出力される文候補の中には、文法的、意味的に正しいものだけでなく、誤ったものも含まれると考えられるが、そのようなものは、用例検索部5における類似度が低くなるため、最終的な音声認識結果とはされない。そして、この場合には、認識部11において、入力単語列が、言語モデルによりかけられる制約により絞り込まれるので、用例検索部5で、類似度の計算対象となる入力単語列の数が、図1における場合と比較して少なくなり、その結果、処理の高速化を図ることができる。

【0062】なお、言語的制約は、言語モデルによる他、例えば有限状態ネットワークなどを用いてかけるようにすることも可能である。

【0063】次に、図7は、本発明を適用した音声認識装置の第3実施例の構成を示している。なお、図中、図1における場合と対応する部分については、同一の符号を付してある。即ち、この音声認識装置は、認識部4および用例検索部5に代えて、認識/用例検索部21が設けられ、さらに言語モデル記憶部12が設けられて、

代えて、拡張単語辞書22が設けられている他は、図1の音声認識装置と同様に構成されている。

【0064】認識／用例検索部21は、認識部4および用例検索部5を一体化したもので、連続音声認識処理と、類似度の計算を並列して行うことができるようになっている。また、拡張単語辞書22は、シソーラス記憶部8に記憶されていたシソーラスと単語辞書6とを一体化したもので、例えば図5に示したシソーラスの単語に、その音韻情報や品詞、その他の必要な情報を対応付けたものが記憶されている。

【0065】以上のように構成される音声認識装置では、認識／用例検索部21において、認識部4における場合と同様に、拡張単語辞書22を参照しながら、音声認識が行われるが、この拡張単語辞書22を参照するときに、音声認識に必要な情報（例えば、単語の音韻情報など）だけでなく、類似度を計算するために必要な情報も読み出されるようになっている。

【0066】従って、その後の類似度計算を行う際には、拡張単語辞書22を参照する必要がないので、処理の高速化を図ることができる。さらに、この場合、拡張単語辞書22は、単語辞書6またはシソーラス記憶部8をそれぞれ独立に実現した場合に比較して、その記憶容量を低減することができる。

【0067】次に、図8は、本発明を適用した音声翻訳装置の一実施例の構成を示している。この音声翻訳装置は、音声認識装置41、翻訳部42、および対訳辞書43から構成されている。音声認識装置41は、図1に示した音声認識装置と同様に構成されている。翻訳部42（翻訳手段）は、音声認識装置41の出力の言語を、対訳辞書43を参照しながら、他の言語に翻訳するようになされている。対訳辞書43は、例えば日本語の単語と、それを英語に訳した英単語（あるいは英単語列）とを対応付けて記憶している。

【0068】なお、音声認識装置41を構成する単語辞書6、用例データベース7、およびシソーラス記憶部8の記憶内容は、例えば日本語に関するものとされている。但し、用例データベース7には、日本語の用例の他、その用例を英訳したものが、対応する日本語の用例に関係（対応）付けられて記憶されている。

【0069】次に、その動作について、図9を参照して説明する。音声翻訳装置では、日本語の音声が入力されると、音声認識装置41において、図1における場合と同様の処理が行われ、これにより、用例検索部5において、最も類似度の高い入力単語列および用例が求められ、その入力単語列および用例が、翻訳部42に出力される。

【0070】ここで、例えば、いま、最も類似度が高くなる入力単語列または用例が、それぞれ（「私」、「学校」、「行く」）または「私は寺に行く」であったとす

校」、「行く」）および用例「私は寺に行く」とともに、その用例に対応付けられている英文の用例としての、例えば「I go to the temple」も翻訳部42に出力される。なお、用例「私は寺に行く」と「I go to the temple」とは、それぞれを構成する、対応する単語どうし（「私」と「I」、「寺」と「temple」、「行く」と「go to」）が、例えば図9（a）に示すように対応付けられている。

【0071】翻訳部42では、（日本語の）用例「私は寺に行く」を構成する単語のうち、入力単語列に対応するもの「私」、「寺」、「行く」が、図9（b）に示すように、入力単語列を構成する単語「私」、「学校」、「行く」にそれぞれ置き換えられ、これにより用例「私は寺に行く」が、「私は学校に行く」に変換される。その結果、この「私は学校に行く」は、「私は寺に行く」と対応付けられていた「I go to the temple」と、図9（c）に示すように対応付けられる。

【0072】その後、翻訳部42は、対訳辞書43を参照し、入力単語列を構成する単語を英単語（あるいは英単語列）に変換する。即ち、いまの場合、入力単語列を構成する単語「私」、「学校」、「行く」が、図9（d）に示すように、「I」、「school」、「go to」にそれぞれ変換される。

【0073】そして、翻訳部42は、図9（e）に示すように、図9（c）に示した「私」に対応する「I」、「学校」に対応する「the temple」、「行く」に対応する「go to」を、図9（d）に示した「私」に対応する「I」、「学校」に対応する「school」、「行く」に対応する「go to」に、それぞれ置き換え、その結果得られる英文「I go to school」を出力する。この翻訳結果は、例えばディスプレイなどで表示され、あるいは音声合成装置に入力されて合成音として出力される。

【0074】従来の音声翻訳装置では、音声認識結果として1つの文が出力された場合に、その文に対して機械翻訳を行うようになされていた。ここで、従来の機械翻訳の手法としては、入力された文を、一旦中間言語に変換し、その中間言語に基づいて、入力文の意味を理解して、その後、目的とする言語に変換するというものが主流であったが、最近では、次のような用例に基づく処理を導入する手法が検討されている。即ち、この手法は、入力されたテキストに類似する用例を検索し、その結果得られた用例を構成する単語を、2言語（入力テキストの言語と翻訳後の言語）間の対訳辞書に基づいて、目的とする言語の単語を置き換え、これにより翻訳結果を得るというもので、この手法によれば、入力文を中間言語に変換する必要がない。

【0075】これに対し、図8の音声翻訳装置では、音声認識において用例に基づく処理を導入しているため、音声認識に利用した用例を、そのまま翻訳に用いること

ぞれ独立に構成して接続した音声翻訳装置に比較して、処理の簡単化および高速化を図ることができる。

【0076】以上、本発明の実施例について説明したが、本発明は、上述した実施例に限定されるものではなく、例えば音声を入力として、その音声に対応した処理を行う装置などに適用可能である。

【0077】なお、本実施例においては、単語類似度を、そのまま積算し、その積算結果を、類似度としたが、この他、類似度は、単語類似度に対し、例えばシソーラスの構造などに対応した重み付けを行い、それを積算したものとするなど可能である。

【0078】また、本実施例では、類似度を求めるのに、シソーラスを利用するようにしたが、類似度を求める方法は、これに限定されるものではない。

【0079】さらに、図8の音声翻訳装置では、日本語から英語への翻訳を行うようにしたが、単語辞書6、用例データベース7、シソーラス記憶部8、および対訳辞書43の登録内容を変更することで、例えば日本語以外の言語を英語に翻訳したり、また日本語を、英語以外の言語に翻訳することなどが可能である。

【0080】また、図8の音声翻訳装置には、図1の音声認識装置と同様に構成される音声認識装置41を設けるようにしたが、音声認識装置41は、この他、例えば図6や図7に示した音声認識装置と同様に構成することなどが可能である。

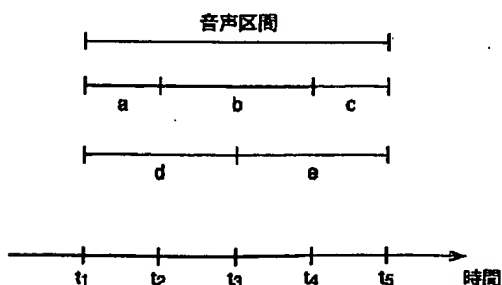
【0081】

【発明の効果】本発明の音声認識装置および音声認識方法によれば、複数の認識結果候補それぞれと、用例それぞれとの類似度が計算され、その類似度に基づいて、音声の認識結果が求められる。従って、文法規則を用いずに、複数の認識結果候補を絞り込んで、音声認識結果を得ることができる。

【0082】本発明の音声翻訳装置によれば、複数の認識結果候補それぞれと、用例それぞれとの類似度が計算され、その類似度に基づいて、音声の認識結果が求められる。そして、その音声認識結果の言語が、他の言語に翻訳される。従って、音声で入力された原語を、容易に翻訳することが可能となる。

【図面の簡単な説明】

【図3】



【図1】本発明を適用した音声認識装置の第1実施例の構成を示すブロック図である。

【図2】図1の用例検索部5の動作を説明するためのフローチャートである。

【図3】図2のステップS1の処理を説明するための図である。

【図4】図2のステップS1の処理を説明するための図である。

【図5】図1のシソーラス記憶部8に記憶されているシソーラスを示す図である。

【図6】本発明を適用した音声認識装置の第2実施例の構成を示すブロック図である。

【図7】本発明を適用した音声認識装置の第3実施例の構成を示すブロック図である。

【図8】本発明を適用した音声翻訳装置の一実施例の構成を示すブロック図である。

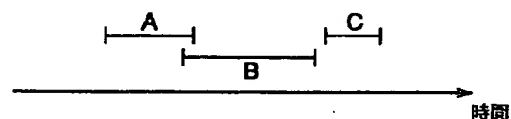
【図9】図8の音声翻訳装置の動作を説明するための図である。

【図10】従来の音声認識装置の一例の構成を示すブロック図である。

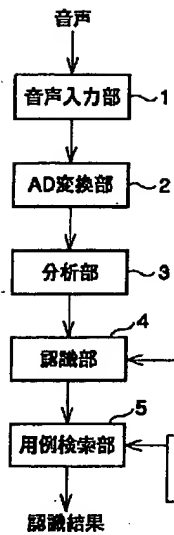
【符号の説明】

- 1 音声入力部
- 2 AD変換部
- 3 分析部
- 4 認識部
- 5 用例検索部
- 6 単語辞書
- 7 用例データベース
- 8 シソーラス記憶部
- 11 認識部
- 12 言語モデル記憶部
- 21 認識/用例検索部
- 22 拡張単語辞書
- 41 音声認識装置
- 42 翻訳部
- 43 対訳辞書
- 51 パーザ部
- 52 単語辞書
- 53 文法規則辞書

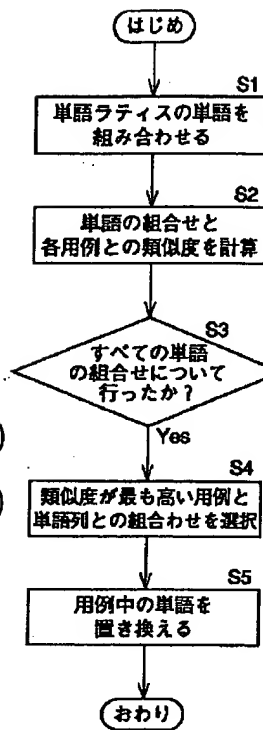
【図4】



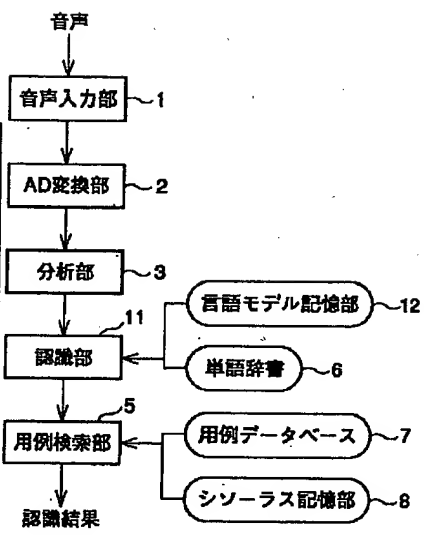
【図1】



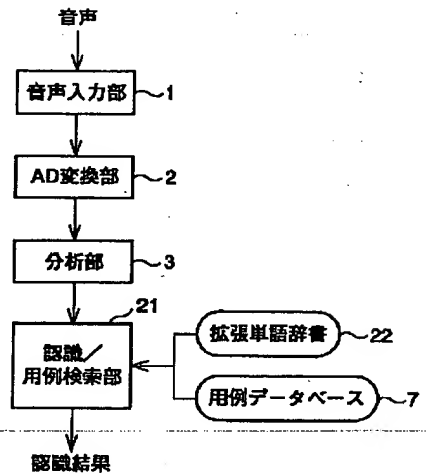
【図2】



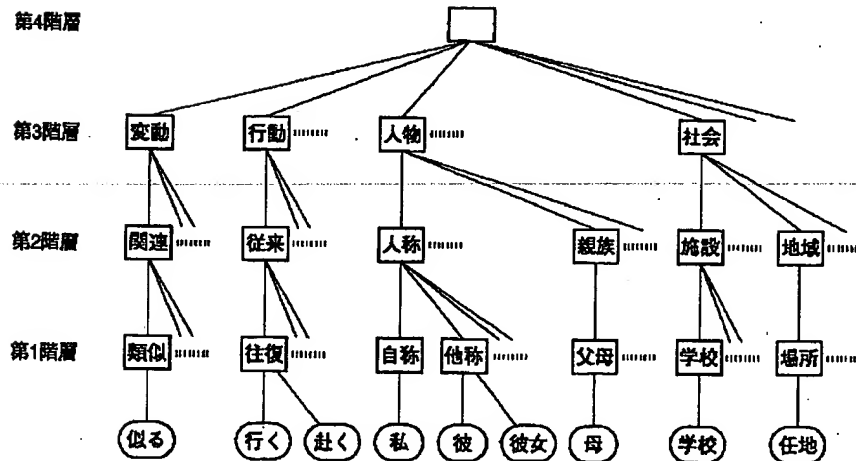
【図6】



【図7】

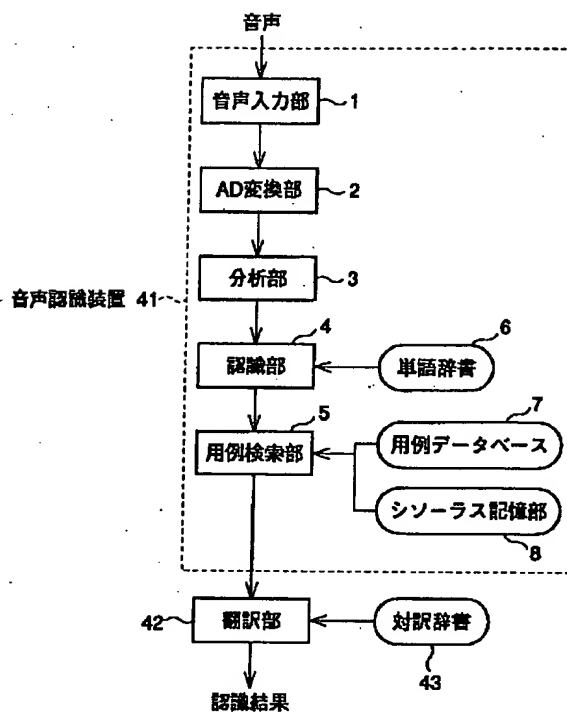


【図5】

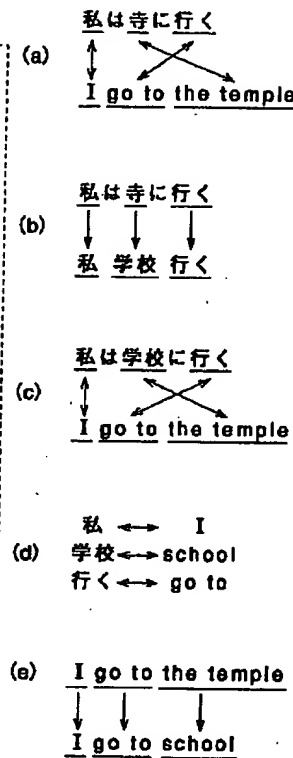


シソーラス

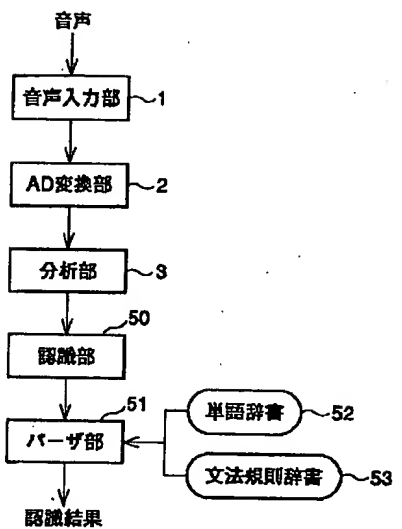
【図8】



【図9】



【図10】



フロントページの続き

(72)発明者 加賀美 徹也
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内
(72)発明者 渡 雅男
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内
(72)発明者 赤羽 誠
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内

(72)発明者 石井 和夫
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内
(72)発明者 田中 幸
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内
(72)発明者 角田 弘史
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内